

EDITORIAL

From systems biology to synthetic biology

Molecular Systems Biology 29 March 2005; doi:10.1038/msb4100007

What is new about systems biology?

If each of us conveys our idiosyncratic view of the elephant, then we, in communion, can imagine its wholeness (Ireland, 1997). An analogous process elucidates Molecular Systems Biology (MSB) and this journal is dedicated to it. The work of a scientist/engineer (seer for short) consists of discovery, modeling, perturbation, and invention. Cycling through these is a key practice of biology. Adding high-throughput analyses (Ideker *et al*, 2001) gives one view of MSB. Systems Engineering (Wiener, 1948; Chestnut, 1967) adds a great deal more. For the next step, we recognize five unalienable rights—to search, check, design, merge, and share (in electronic media). Exercise of these is exemplary in the realms of genetics, crystallography, and sequencing. Few who use Google or Blast will doubt that text-searching is a killer-application. With computer searches, we can check if a discovery or design is truly new or alignable with other facts. If alignable, merging can produce a more comprehensive map of the elephant. Without these five rights, seers fall to politics, opinions, and fads. But even the most meager search presupposes a means to share and to decide on alignments. Journals and granting agencies encourage sharing promptly and unambiguously via accession numbers. We generally share models, not raw data. Aligned protein models are far beyond raw DNA electrophoresis data; aligned 3D structures are far from diffraction spots; and genetic linkages are far from DNA-polymorphism-chip intensities. For each of these, we have useful (and jargon-laden) metrics of goodness of fit and completion (Selinger *et al*, 2003), for example, Blast *E*-values, *R*-factors/RMSDs, and LOD scores, respectively.

How can the rest of biology achieve this enviable state?

The current state of the art for many biologists is to share images of cells and gels and then describe models with circles and arrows. While these data and models are technically machine-readable in e-journals, they are not sufficient for the search/check/design/merge/share tools that we covet. As the comprehensiveness of genome sequencing begins to extend to functional genomics (also known as 'omics'), quantitating RNA, protein, and metabolic domains, we must confess our inability to keep all of this in our heads well enough to evaluate new discoveries. It is an important exercise for each of us, as we contemplate publishing our latest discovery to ask how we can make this more accessible for checking and merging with other discoveries. Tools for modeling and *de facto* standards, for example, SBML/BioSPICE (Kumar and Feidler, 2003), are

emerging. We cannot let the future-perfect be the enemy of the current-good, as we will learn best by doing.

How will we evaluate our progress?

As we prepare our manuscripts, we ask, 'Could someone with a computer reproduce our logic in getting from our data to our model?' To what extent can they do this without reading between the lines or emailing us for clarification? Are all of the needed data and programs available online? For example, say, instead of merely claiming that an image of GFP in transfected cells can only be used qualitatively, we should give a supplementary set of images, tables of quantitative measures (however inconclusive relative to the accepted eyeballing method), and the software used. This approach is more open and will challenge the math-jocks to see if they can align eyeballing and automatic output better. As the paper transitions to causal analyses, adding an arrow between two circles gives the odds that the arrow should be added given the data above. The vague Ockam's razor will sharpen into the rigors of multiple hypothesis testing. Data sets constructed from known connectivities and parameters will allow us to test and compare algorithms for inferring network topology and parameters even in the face of overlapping cycles, experimental errors, and large dynamic ranges.

To aid merging with other models, machine-readable descriptions of experimental composition (genotype and environment) is crucial (Aach *et al*, 2000; Selinger *et al*, 2003).

Initially, researchers try to keep genotype and environment constant via standard 'model systems', but our ability to evaluate and exploit merged models over increasing experimental distances is growing rapidly. This is evident as we take hints from yeast, worms, and flies and apply them to human experimental systems. This process was accelerated in the realm of sequence comparisons by the advent of machine-readable sequences and, hopefully, something analogous will happen soon with biology via 'systems biology'. I sincerely hope that this journal will be at the cutting edge of this revolution. Just as molecular biology was once a tiny fraction of biology, but now affects all corners, so, too, systems biology is likely to be embraced by all of biology as quickly as tools can be created and distributed. Early on, we will note that many systems are underdetermined, that is, the number of adjustable parameters is more than the number of experimental data. This can be fixed by adding constraints (Price *et al*, 2004), separating out subsystems (genetically or biochemically), and of course, developing new technological sources of data. If probabilistic constraints exist, they should be applied, at least, as a means for discovering the limitations of the

constraints. A terrific source of constraints is the optimality of many biological subsystems for their evolved tasks, and deviations from optimality are expected in certain cases (like mutants) and are informative (Segre *et al.*, 2002).

The payoff for systems biology research is not merely abstract mathematical understanding, but empowerment to design new and improved biological functions via ‘synthetic biology’ (Silver and Way, 2004). Combinations of well-characterized biological parts to create synthetic wholes not only drives toward applications faster but also finesses past the underdetermination and crosstalking nonmodularity of natural systems. With the advent of facile synthesis and reusable modules, the evolutionary bricolage can be studied or avoided as needed. We routinely ‘program’ ourselves with small molecules, proteins, nucleic acids (e.g. vaccines), and stem cells (e.g. hematopoietic). With the rapidly increasing availability of personal medical omics data, along with databases of correlation and causality, physicians and researchers alike will need software to help make and explain complex probabilistic decisions to prioritize diagnostics, preventative lifestyles, and therapies. We hope that systems biology, in general, and this journal, in particular, will play central roles in expanding and deploying these revolutionary concepts tools, data, and models.

References

- Aach J, Rindone W, Church GM. (2000) Systematic management and analysis of yeast gene expression data. *Genome Res* **10**: 431–445
- Chestnut H. (1967) *Systems Engineering Methods*. New York: Wiley
- Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* **2**: 343–372
- Ireland JD (1997) Translation of 1st century BC. *Tittha Sutta (Udana VI.4) Parable of the Blind Men and the Elephant*. Buddhist Publication Society
- Kumar SP, Feidler JC (2003) BioSPICE: A computational infrastructure for integrative biology. *OMICS: J Integrative Biol* **7**: 225–225
- Price ND, Reed JL, Pálsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2**: 886–897
- Segre D, Vitkup D, Church GM. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99**: 15112–15117
- Selinger D, Wright M, Church GM. (2003) On the complete determination of biological systems. *Trends Biotechnol* **21**: 251–254
- Silver P, Way J. (2004) Cells by design. *Scientist* **18**: 30–31
- Wiener N (1948) *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: The MIT Press

George M Church

Department of Genetics, Harvard Medical School